AFRL-HE-AZ-TR-2006-0012

Evaluating Competency-based Performance: A Probabilistic Approach

Thomas F. Carolan

Micro Analysis & Design Inc
4538 Manitou Drive
Okemos MI 48864

Ira Schurig, Captain, USAF
Winston Bennett Jr

Air Force Research Laboratory
Warfighter Readiness Research Division
6030 South Kent Street
Mesa AZ 85212-6061

December 2006

Final Report for September 2003 to February 2006

Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Readiness Research Division

NOTICES

AFRL-HE-AZ-TR-2006-0012 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

// Signed //

WINSTON BENNETT JR
Laboratory Task Monitor

// Signed //

HERBERT H. BELL
Technical Advisor

// Signed //

DANIEL R. WALKER, Colonel, USAF
Chief, Warfighter Readiness Research Division
Air Force Research Laboratory

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
15-12-2006

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
September 2003 to February 2006

4. TITLE AND SUBTITLE
Evaluating Competency-based Performance: A Probabilistic Approach

5a. CONTRACT NUMBER
F41624-97-D-5000 TO 34

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
63231F

6. AUTHOR(S)
Thomas F. Carolan
*Ira Schurig
*Winston Bennett Jr

5d. PROJECT NUMBER
4924

5e. TASK NUMBER
AS

5f. WORK UNIT NUMBER
02

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Micro Analysis & Design Inc
4538 Manitou Drive
Okemos MI 48864

*Air Force Research Laboratory
Warfighter Readiness Research Division
6030 South Kent Street
Mesa AZ 85212-6061

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Readiness Research Division
6030 South Kent Street
Mesa AZ 85212-6061

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL; AFRL/HEA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-HE-AZ-TR-2006-0012

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
This paper was presented at the 2003 Behavior Representation in Modeling & Simulation (BRIMS) Conference, held 12-15 May 03, in Mesa AZ.

14. ABSTRACT
This paper describes part of an ongoing research program designed to integrate both objective and observer-provided data to develop comprehensive tools for assessing and diagnosing pilot performance in complex and dynamic training and rehearsal environments. The goal is to provide a probabilistic capability to assess pilot knowledge and skill competencies and to provide results to instructors for their use in the remediation of performance and the identification of "gaps" that remain. The development process and efforts to date are reported.

15. SUBJECT TERMS
Bayesian networks; Competency-based performance; Distributed Mission Operations; DMO; Instructional technology; Mission Essential Competency; Pilot performance; Probabilistic approach; Skill competencies; Training systems technology

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Dr Winston Bennett Jr.
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED	12	**19b. TELEPHONE NUMBER** *(include area code)*

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

This page intentionally left blank.

PREFACE

This research effort was conducted for the Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter Training Research Division, Mesa AZ, under Air Force contract F41624-97-D-5000, Task Order No. 34, and Work Unit 4924-AS-02, Warfighter Readiness Assessment and Performance Measurement Tracking System.

The work documented in this report was presented at the 2003 Behavior Representation in Modeling and Simulation Conference held 12-15 May 2003 at the Mountain Shadows Resort in Mesa AZ, and was published in the proceedings..

Note that the current organization name is the Air Force Research Laboratory, Human Effectiveness Directorate, Warfighter <u>Readiness</u> Research Division.

This page intentionally left blank.

Evaluating Competency-Based Performance: A Probabilistic Approach

Thomas F. Carolan., Ph.D.
Micro Analysis & Design, Inc.
4538 Manitou Drive
Okemos, Michigan 48864
517-347-6117
tcarolan@maad.com

Capt. Ira Schurig
Winston Bennett, Jr., Ph.D.
Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Training Research Division
6030 South Kent Street
Mesa, Arizona 85212-6061

Keywords:
Performance Measurement, Mission Essential Competencies, Bayesian networks

Abstract: *This paper describes part of an ongoing research program designed to integrate both objective and observer-provided data to develop comprehensive tools for assessing and diagnosing pilot performance in complex and dynamic training and rehearsal environments. The goal is to provide a probabilistic capability to assess pilot knowledge and skill competencies and to provide results to instructors for their use in the remediation of performance and the identification of "gaps" that remain. The development process and efforts to date will be reported.*

1. Introduction

Researchers have found it difficult to create instructionally effective simulations because the state of the art in instructional technology for simulation is weak (O'Neil & Andrews, 2000). Instructional training research using virtual environments is still relatively new and has not provided the significant research base needed to make training program design decisions. Many instructional strategies are based on cognitive, educational, and learning theories. The focus of instructional strategy research has been on how learners acquire knowledge and then linking performance to specific instructional principles. Although we now have an extensive research base on how learners acquire knowledge, the more difficult and relevant issue is the quantity and quality of practice necessary to achieve effective training performance.

The Air Force's Distributed Mission Training (DMT) program as an exemplar of Advanced Distributed Learning (ADL), is a major advance in ground-based training that will allow pilots and other warfighters to train for complex, multi-player combat operations. Researchers from the Air Force Research Laboratory, Warfighter Training Research Division (AFRL/HEA) are investigating strategies for using DMT to augment advanced flying training in operational units. The principled design of DMT scenarios represents a middle ground between single-ship simulator training and large-force exercises. In single-ship simulator training such as learning to respond to in-flight emergencies, an instructor introduces an emergency such as an engine malfunction and then waits for the student to respond. Events are highly scripted and the instructor can readily evaluate good vs poor performance. In contrast, large-force exercises are much less scripted at the level of individual pilots. Evaluators will know where and when forces will engage but will have only limited control over each pilot's experience. Tying scenario events to mission essential competencies, and by reference, to training objectives and specific trainee behaviors, provides the basis for instructor evaluations of team or individual performance. The instructor knows for any given moment in a scenario what competencies are being tapped, what objectives are being trained, what trigger events are about to occur, and what behaviors are critical to mission success (Bennett & Crane, 2002).

Mission Essential Competencies (MECs) for aircrew training performance (Colegrove & Alliger, 2002) identify the critical knowledge and skills necessary for successful air combat, and provide a framework for measuring knowledge and skill competencies. A primary goal of performance measurement is to identify strengths and weaknesses in the knowledge and skills necessary for successful air combat so that training can be focused on addressing identified deficiencies.

This paper describes an attempt to evaluate changes in Aircrew knowledge and skill competencies that develop over DMT training sessions. The approach is designed to use both automated and observer generated performance data as evidence for the strength or weakness of particular competencies. The resulting profiles would provide information to support adaptive training through the selection of scenario elements for future training.

1.1 Competencies required for successful performance

Mission Essential Competencies are described as higher-order individual, team, and inter-team competencies that a fully prepared pilot, crew or flight requires for successful mission completion (Colegrove and Alliger, 2002). Mission Essential Competencies are demonstrated in the context of an actual mission or high-fidelity simulated mission. For example, "Intercepts and targets factor groups" is one of the MECs for Air Superiority.

Mission Essential Competency development involves different levels of detail (Colegrove and Alliger, 2002). Mission Essential Competencies include a more detailed decomposition of competencies that more fully describes each Mission Essential Competencies. Personnel that exhibit high levels of proficiency in a Mission Essential Competency are also proficient in a series of sub-competencies that support the Mission Essential Competency. These supporting competencies are sets of high-level skills. Situational awareness, communication, and decision-making are all examples of supporting competencies. Some supporting competencies are applicable across all Mission Essential Competencies, and others are applicable for only one or two Mission Essential Competencies. Supporting competencies can be broken down even further into knowledge and skills. A variety of knowledge and skill requirements are necessary in attaining a supporting competency. Example Knowledge requirements include: "Understands threats, their capabilities, and their tactics", "Knows criteria for commit decision," and "Understands formation standards". Examples of skill requirements include: "Builds picture", "Controls intercept geometry", and "Selects tactic".

1.2 Competency evaluation goals

A significant requirement for continuous improvement and maintenance of proficiency is an evaluation process that can identify proficiency levels on core competencies and use this information to focus training to challenge appropriate competencies and maximize learning. The primary goal of the project is a semi-automated process that provides evaluative information about knowledge and skill competencies based on observed performance during DMT exercises. A semi-automated evaluation process combines objective performance information automatically generated using training simulation data files and both objective and subjective performance information generated by instructor/observers and perhaps by the pilots themselves.

Performance evaluation data derived from objective simulation-based measures and observation based measures provide the basis for assessment of the knowledge and skills that support each MEC (e.g., Schreiber, MacMillan, Carolan & Sidor (2002). Assessing knowledge and skill proficiencies based on performance data can be thought of as assigning "credit or blame" to a knowledge or skill element or combination of elements for observed performance deficiencies. The goal is to develop individual and team competency profiles based on performance over a single DMT exercise and a series of DMT exercises. The competency profiles can then be used to track progress and tailor exercises based on individual and team mastery or lack of mastery of specific competency areas (Bennett, Schreiber & Andrews, 2002).

1.3 Capturing objective performance data

Recent research and development at the Air Force Research Laboratory in Mesa, AZ has resulted in a proof-of-concept automated distributed performance effectiveness and evaluation tracking system (PETS). The need was to create an automated objective measurement tool that would assess both higher- and lower-level F-16 air combat MECs in a DMT environment.

Interactive, distributed simulation environments such as DMT typically adhere to Distributive Interactive Simulation (DIS) or High Level Architecture (HLA) standards. With these standards, data is passed on a network. A performance effectiveness system could therefore reside on the network and listen to the network traffic, collect appropriate variables at any rate specified, input/output variables through measurement algorithms as necessary, and output the data in several data formats, either for feedback or for statistical analysis purposes. However, the data in and of themselves do not have the diagnostic or predictive qualities necessary for evaluating the performance, proficiency and or mastery of trainees without additional modeling and validation (Bennett, Schreiber & Andrews, 2002).

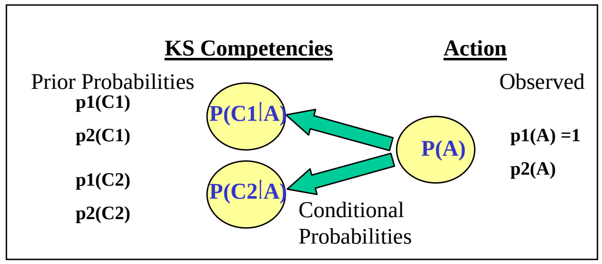
2. Issues

In the DMT environment observed performance will provide only limited information about the knowledge and skill competencies underlying performance. In addition the accessibility of performance data to support automated assessment while dramatically improving is still limited in the short term. The observed information is typically incomplete, the number of competencies to be assessed is quite high compared to the amount of performance data available, and there are typically many different paths to a particular performance outcome. Probabilistic reasoning provides a useful methodology for developing assessments in environments where the assessment is performed under conditions of uncertainty.

3. Overview of Bayesian Approach

Bayesian networks allow probability-based inference from observable variables (e.g., performance) to hypothesized non-observable variables (e.g., knowledge and skill competencies). Bayesian networks involve mathematical methods that permit reasoning under conditions of uncertainty based on Bayes theorem. The Bayesian belief network, or Bayesian network, methodology is a relatively recent development for simplifying the computationally complex Bayesian reasoning process (Charniak, 1991). Using Bayesian networks for diagnostic assessment and modeling trainee competencies is a developing area of research (e.g., Nichols, Chipman & Brenan, 1995). Bayesian network technology has been applied to diagnostic assessment in computer-based tutoring systems in academic and applied research environments (e.g., Gitomer, Steinberg & Mislevy, 1995; Martin & VanLehn, 1995; Mislevy, 1995). In these environments, probabilistic reasoning is often used as a mechanism to diagnose the knowledge, skills, and/or strategies that are used in solving a particular problem, making a decision or performing an action.

A Bayesian network is a graph structure where the nodes represent variables with two or more possible values (e.g., true, false). Links represent conditional probability relations between the values of the variables. The nodes generally represent one of two types of variables, observable events or actions and situations or conditions to be assessed based on those events. The observed events provide evidence for the values of the related non-observed variables. Figure 1 illustrates the basic components as applied to a generic proficiency example. The assessment of proficiency is to be updated based on the observed action.



$$P(C|A) = \frac{P(A|C)P(C)}{P(A)}$$

Figure 1. Basic components of a Bayesian network

The process can be roughly described as follows. Observable actions are defined with an expected probability distribution, P(A). Competencies are defined with estimated proficiency levels characterized by a probability distribution, P(C). Conditional relationships are defined between competencies and performance, quantifying the likelihood of an action occurring given distributions of competency values, P(A|C). Actions are observed and entered as evidence, e.g., P1(A) = 1. Joint probabilities relating competencies and actions are updated. Posterior Probabilities for competencies given observed performance, P(C|A), are returned.

Applying a Bayesian network approach requires three sources of information. The first is information about the potential values of the competency variable to be evaluated and the likelihood that the competency is in each of those potential states prior to the observed event. For example, a competency variable might be represented as having two values, high and low. The prior probabilities might be p = .4 and p = .6 respectively. These values could be based on prior performance history or knowledge of the trainee population. The second source of information is the relation between observed events and the variables to be evaluated. The likelihood (conditional probability) of observing each value of the action variable given the possible states of the competency variables is quantified. For example, the probability that an individual or team with a high proficiency rating for "Understanding of formation standards" would deviate from appropriate formation parameters would be quantified, perhaps at the p=.4 level to reflect the role of other factors and competencies. A low proficiency team might be expected to deviate from formation standards with a much higher probability. The source of the conditional probability estimates can come from empirical data, if it is available or, more likely from expert judgements. The third source of information is the observed actions. When actions are observed the prior and conditional probabilities are used to update the variables, producing posterior probability values for the competency variable. A significant deviation from formation standards would reduce the probability that the individual in question has an acceptable understanding of formation standards.

4. Application Approach

Our approach to implementing a probabilistic KS competency evaluation strategy for DMT can be thought of as a successive approximation approach. Many of the thirty-two knowledge and skill elements have an impact on performance elements across all phases of a mission. Most performance requirements involve a range of knowledge and skill elements combining to produce effective results in each MEC area. This first approximation approach to developing student competency profiles starts by identifying conditional relations between specific performance requirements and relatively high level knowledge and skill elements. At this level, rather than modeling how a knowledge or skill element might impact performance on a particular task, we are identifying only which knowledge and skill elements are required for effective performance and the relative impact of each on success or failure of the task. This first approximation approach provides some benefits. First, it allows us to develop a method for automating the construction of competency networks. Second, it allows us to limit the depth of the competency networks. Third it allows us to develop networks for each MEC and add performance measures, as they become available. Of course there are clear limitations. First, this approach does not consider any other sources of evidence for the relative role of specific knowledge and skill elements on observed performance. Second, this approach does not consider the influence of previous actions on performance. Third, this approach does not provide a strategy for diagnosing specific performance measures in terms of knowledge/skill competencies.

However, these limitations can be addressed with additional effort. This first approximation level of assessment may provide a useful basis for characterizing Air Superiority knowledge and skill competencies as a basis for selecting training interventions. It should also provide useful data on changes in specific competencies as a result of DMT exercises.

5. Performance and Competencies Requirements Analysis

A team of six expert pilot trainers is involved in an intensive workshop approach to the task and performance analysis process. Detailed task and performance requirements are developed for each of the MECs. The decomposition provided the information needed to support performance evaluation. Performance measures were identified for each item and performance standards were defined. To support performance evaluation, information about when to measure (triggers), who to measure, and what to measure was specified for each item. The source of the evaluation data was identified as either simulation based or observer based and an assessment was made as to the likely availability of performance data to support each measure. Supporting competencies and knowledge and skill elements are assigned to the MEC tasks and the relative impact of each knowledge and skill element on task performance is rated.

5.1 Competency Requirements Analysis

For each performance requirement, the knowledge and skill competencies required for successfully achieving performance criteria are identified. The knowledge and skill items are then assigned weights, from 1 to 5. The weight values are anchored by a descriptive definition that indicates the importance of the knowledge or skill to successful performance and by a probability value. The probability value is best defined as the likelihood that, if the observed performance measure indicates substandard performance, a deficiency in the particular knowledge or skill has some causal responsibility. It is an estimate of the impact a deficiency in the knowledge or skill would have on task performance. The relation between weights and expected performance are as follows:

Weight 1: p = .1
Weight 2: p = .25
Weight 3: p = .5
Weight 4: p = .75
Weight 5: p = .9

These probabilities can be interpreted as the probability that, if a knowledge or skill competency with a given weight is missing or weak, the performance requirement for the associated task will not be met with the given probability. For example, if a competency with a weight of 5 (it is essential for completion of the task) is the only one missing, there is a .9 probability the performance requirement will not be met.

6. Implementation: Performance Evaluation

To simplify and structure the automated performance evaluation problem, the evaluation approach is to compare observed performance to a standard solution. Performance that deviates from the standard or preferred solution is identified for evaluation and further discussions during debrief. The evaluation process is implemented as a Performance and Competency Evaluation Support tool (PACES).

The PACES database is populated with the performance requirements analysis data. The data is entered as a task hierarchy and the measures associated with each task. A degree of authorability is provided. Tasks and measures can be added, deleted or edited. An existing measure can be added to the task and its properties modified or a new measure can be added to the list. Associated with each measure are the properties required to support evaluation of performance and evaluation of mission essential competency elements. Measure properties required for performance evaluation include the triggers, standards and computational formula. Start and stop triggers are defined for each measure. Triggers can be based on data file variables or on user defined variables or a combination of both. User defined variables refers to variables that must be identified by an observer (or some other agent) in order to trigger a measurement start or stop. The measure standards property dialogue provides a way to manually enter the performance standards, the position to whom it applies, the units of measure, the permissible deviation, and a link to a reference or reference document. An "import standards" function allows the briefed standards to be read in from a file, and a "select standards" option allows a particular set of standards to be used when running an engagement analysis.

6.1 Observer-based measures

While the emphasis is on using objective performance measures to evaluate competencies, observer-based objective and subjective performance measures are also required. The performance measures identified during analysis include objective measures that can be evaluated using performance data, objective data that cannot now be measured or require an instructor/observer component and subjective measures that can only be captured by an instructor/observer during the exercise or during debrief.

The database structures, software functions and user interface to support authoring observer-based performance measures are implemented within PACES. PACES includes a tool to author "manual" performance measures for each performance element, an assessment interface for the observer to collect or evaluate performance data, and the functionality to evaluate and integrate observer and simulator data for evaluation. The authoring tool provides the option to define various types of *measure data types* using a range of GUI objects and provide behavioral anchors or just generalized evaluation instructions for each measure value.

7. Implementation: Competencies Evaluation

A benefit of this strategy is that competency networks can be generated automatically once the KS weights are assigned to a measure. The capability to automatically construct the competency networks makes it easier to test and refine the networks, add new competencies, and build new networks for new missions.

The KS competency model structure can be different for each analysis since there will be multiple instances of many measures for each participant depending on factors particular to the exercise, such as number of threats, number of attacks, etc. While the competency network and weights will be the same, the number of measure instances assigned to each competency will be different, requiring the generation of a new assessment network for each analysis. For each analysis of scenario data, PACES constructs competency networks for each MEC and each participant using information about the competencies to be evaluated, the weights relating competencies to performance, the performance data associated with each competency and the current competency profiles of the exercise participants.

The weights relating KS competencies to performance measures are converted to a set of conditional probabilities for the success of each measured performance variable given each possible combination of relevant KS competencies. The algorithm emphasizes the probability of failure to achieve the performance standard given a weak competency. For each combination of competencies, the conditional probability is generated by applying the weights in descending order to the remaining probability of a successful outcome, reducing it by an amount proportional to the probability value assigned to the weight value. The Bayesian does not have any information about how the probability for each KS combination was computed. Therefore, it treats each weight of the same value in the same way.

Once each network is constructed, the performance evaluation values are used to update the evidence variables. All the competency variables are then updated based on the performance evidence, generating posterior probability values from the competency prior probabilities, the evidence values and the joint probabilities relating the two.

7.1 Defining prior probabilities

The analysis user interface provides an option to run each post exercise analysis using either the current competency profile for each participant or "default" priors (currently set at p =. 5) that do not reflect participant history and expertise. The two options provide different information and each can be useful. Updating individual competency profiles provides a means to track proficiency or performance readiness on each of the competencies and can be used as an indicator of DMT training effectiveness. A record of the competency profile after each exercise is maintained in the database to be used to set the prior probabilities (the expected level of competency on each KS for each individual) for the next scenario. This approach allows the history of the individual's performance to be considered in evaluating the current performance.

Running the competency analyses without prior performance history provides a way to collect baseline competency data for each exercise. In this approach the individual history of performance is not considered in the evaluation of the current exercise.

7.2 Accessing analysis results

The performance and competency analyses are run as separately. Running the performance analysis calculates each measure and compares it to the performance standard. The output of the analysis is a list of the deviations from the performance standard for each measure. For each deviation instance, the start time, total time, average deviation and maximum deviation are captured. For each analysis, the user selects the appropriate performance standards file, whether to run it for all MECs or only one MEC, and whether to run it for all participants or only one participant. All is the default case

The performance measure evaluations provide the input to the competency network. Running the competency analysis sends the performance measure evaluation data for the engagement scenario to the competency models. The output of the competency analysis consists of a score between 0 and 1 for each competency for each participant under each MEC based on performance on the particular exercise. These updated competency profiles (posterior probabilities) are stored and available to be used as the initial values of the competency variables (prior probabilities) for the next exercise.

PACES provides two levels of performance results. The top level is a color-coded list of measures. Those measures where performance deviated from the standard parameter values are indicated in red. The next level of results provides the performance detail (time of each deviation, length of deviation, average deviation, maximum deviation) for those measures selected. The display for viewing competency results consists of three levels. The top level provides the Knowledge and Skill proficiency values. The values are between 0-1. Scores are color-coded using the stoplight metaphor with user defined threshold values. The values can be interpreted as measures of the strength of each competency based on performance in the engagement (and the distribution and weighting of competencies over the performance measures). The measures that contributed to each knowledge and skill competency value can be viewed. The third level consists of the performance detail for each measure.

8. Evaluation Process

Plans for evaluation and refinement of the competency networks involve comparing the output of the networks with the ratings of human evaluators for a number of engagements. Another approach under consideration is developing performance models for specific tasks and comparing the added value of these performance models, as a means to improve the ability to differentiate between competencies. Test scenarios have been flown specifically for evaluation and development of the competency networks. Preliminary results should be available before the final paper draft is due in April.

9. References

[1] Bennett, W., Jr., & Crane, P. (2002). The deliberate application of principles of learning and training strategies within DMT. Paper presented at the NATO SAS-038 Working Group Meeting, Brussels, Belgium.

[2] Bennett, W., Jr., Schreiber, B.T., & Andrews, D.H. (2002). Developing competency-based methods for near-real-time air combat problem solving assessment. Computers in Human Behavior, 18 (773-782).

[3] Charniak, E. (1991) Bayesian networks without tears. AI Magazine, Winter (50-63).

[4] Colegrove, C.M., & Alliger G.M. (2002). Mission essential competencies: Defining combat mission requirements in a novel way. Paper presented at the NATO SAS-038 Working Group Meeting, Brussels, Belgium.

[5] Gitomer, D.H., Steinberg L.S. & Mislevy R.J. (1995). Diagnostic assessment of troubleshooting skill in an Intelligent Tutoring System. In Nichols, P. D., Chipman, S. F., & Brennan, R. L. (Eds.). *Cognitively*

diagnostic assessment. Hillsdale, NJ: Lawrence Erlbaum Associates.

[6] Martin, J. & VanLehn K. (1995). A Bayesian approach to cognitive assessment. In Nichols, P. D., Chipman, S. F., & Brennan, R. L. (Eds.). *Cognitively diagnostic assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates.

[7] Mislevy, R. J. (1995). Probability-based inference in cognitive diagnosis. In Nichols, P. D., Chipman, S. F., & Brennan, R. L. (Eds.). *Cognitively diagnostic assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates.

[8] Nichols, P. D., Chipman, S. F., & Brennan, R. L. (Eds.). (1995). *Cognitively diagnostic assessment*. Hillsdale, NJ: Lawrence Erlbaum Associates.

[9] O'Neil, H. F., & Andrews, D. (Eds.). (2000). Aircrew training and assessment. Hillsdale, NJ: Lawrence Erlbaum Associates.

[10]Schreiber, B.T., MacMillan, J., Carolan, T., & Sidor, G. (2002). Evaluating the effectiveness of Distributed Mission Training using "traditional" and innovative metrics of success. Paper presented at the NATO SAS-038 Working Group Meeting, Brussels, Belgium

Author Biographies

THOMAS F. CAROLAN is a Senior Research Psychologist with Micro Analysis & Design, Inc. Boulder CO. He received his Ph.D. in Experimental Psychology from the University of Connecticut. Dr. Carolan is involved in research and development related to training systems, performance measurement, and cognitive and performance modeling.

CAPT. IRA A. SCHURIG is a Behavioral Scientist in the Training Research Branch, Warfighter Training Research Division, Mesa Research Site, Air Force Research Laboratory, Williams Gateway Airport, Mesa, AZ. He received his master's degree in Industrial/Organizational Psychology from St. Mary's University in 2002. He is involved in distributed mission planning, briefing, and debriefing research and development at AFMC's premier warfighter training research organization, integrating R&D efforts to ensure optimum solutions to real world training requirements for customers in the DoD, industry, and contracting.

WINSTON BENNETT, JR. is a Senior Research Psychologist with the Air Force Research Laboratory Human Effectiveness Directorate, Warfighter Training Research Division, located at Williams-Gateway Airport, Mesa AZ. He is the team leader for training systems technology and performance assessment research and development. He received his Ph.D. in Industrial Organizational Psychology from Texas A&M University in 1995. Dr. Bennett has published over 75 research articles, book chapters and technical reports in the Human Resources arena. He is actively involved in research related to performance evaluation, personnel assessment, training requirements identification, and quantifying the impact of organizational interventions--such as interactive, high-fidelity immersive simulation environments and job redesign/restructuring and training systems--in individual, team, and organizational effectiveness.